Active Learning for Improving the Classification Accuracy of Artificial Neural Networks(ニューラルネットワークの分類能力向上のための能動的学習)

氏名（本籍）	滝沢寛之（新潟県）
学位の種類	博士（情報科学）
学位記番号	情博第128号
学位授与年月日	平成11年9月9日
学位授与の要件	学位規則第4条第1項該当
研究科，専攻	東北大学大学院情報科学研究科（博士課程）情報基礎科学専攻
学位論文題目	Active Learning for Improving the Classification Accuracy of Artificial Neural Networks （ニューラルネットワークの分類能力向上のための能動的学習）
論文審査委員	（主査） 東北大学教授 中村維男　東北大学教授 堀口剛 東北大学教授 山本光璋　東北大学教授 猪岡光 東北大学助教授 小林広明

論文内容要旨

A main application of artificial neural networks(ANN) is pattern recognition. Usually, a pattern recognition problem can be decomposed into subproblems, such as feature extraction and classification. This thesis focuses only on application of ANNs to the classification problems. The most popular ANN model widely applied to those problems is called *a multilayer perceptron* (MLP). As an MLP is theoretically an universal approximator, it is expected to behave as, in a sense, an optimal classifier for any classification problem.

One of the difficulties in constructing a classifier by an MLP is that an MLP requires quite a large number of training data to improve its classification accuracy. Here, each training datum is a pair of an input vector and its corresponding correct output vector. Correct output vectors used in training data are generally provided by an external supervisor so called *an oracle*. An oracle could be a computationally intensive simulator, a database obtained by expensive experiments, or a human expert. Therefore, it could become costly to construct a training data set that includes the enough number of training data for MLPs to achieve sufficiently accurate classification.

Active learning is a general term for learning scenarios where the classification accuracy of an MLP is improved by selecting input vectors according to a certain strategy and using those vectors as new data. Active learning usually allows MLPs to achieve accurate classification with fewer training data than conventional training data generation called *passive learning*.

The objective of this thesis is to establish new active learning to more effectively construct an appropriate training data set for accurate classification at lower computational cost. For this purpose, this thesis proposes a novel active learning strategy. The proposed strategy attempts to generate only necessary training data for accurate classification, and constructs an appropriate training data set with less redundancy.

If the proposed strategy turns out to be generally applicable, several benefits are immediately apparent for a wide range of situations, because an MLP can behave as a classifier with a small training data set. First, it leads up to a reduction in the computational cost of constructing a training data set. This reduction is important because the acquisition of each training datum is very costly in many situations, e.g., an oracle may be a human expert. Second, it is possible to save the storage capacity for a training data set required by MLPs for accurate classification. Third, as the computational cost of training MLPs depends greatly upon the number of training data used in the training task, the processing time required for the task would be reduced even in the case where an oracle is not computationally intensive.

In order to realize the training data generation based on the proposed strategy, two active learning algorithms are proposed in this thesis. In addition, some techniques for acceleration of the proposed algorithms are discussed to reduce their computational cost.

This thesis consists of five chapters as outlined below.

1 Introduction

This chapter is the introduction, and describes the objective and background of this thesis as noted above.

2 Active learning

Most of the existing active learning algorithms generate training data that are considered *useful* for improving the classification accuracy of MLPs when generated. However, those useful training data cannot always contribute to the achievement of accurate classification, because they might become *redundant* in the future by generating other useful training data. If there is a method to avoid generating those temporarily useful training data, an MLP can achieve accurate classification with fewer training data in comparison with those existing algorithms.

For classification purposes, an MLP has to estimate the correct boundary among the classes of a classification problem being learnt, so called *a decision boundary*. An MLP can behave as an optimal classifier if it can correctly estimate the whole decision boundary. Generally, a training data set with less redundancy includes more training data lying nearby the decision boundary and fewer training data lying far from the boundary. This is because training data lying far from the decision boundary tend to become redundant in the future, even if they are considered useful when generated.

A representation of the decision boundary formed by an MLP is called *a classification boundary*. To construct a training data set with less redundancy, this chapter defines the reliability of the classification boundary as follows:

> A part of the classification boundary is considered reliable if the part is expected to locate nearby the decision boundary. On the other hand, a part of the boundary that has a possibility to locate far from the decision boundary is considered unreliable.

As only unreliable parts of the classification boundary can locate far from the correct decision boundary, additional training data that can eliminate the unreliable parts are useful for improving the classification accuracy of an MLP. However, such training data could become redundant in the future, because they usually locate far from the decision boundary. On the other hand, although training data lying upon the reliable classification boundary cannot drastically improve the classification accuracy, they are necessary for the MLP to achieve accurate classification.

Based on the above observations, this chapter proposes a novel active learning strategy named *active learning based on existing training examples* (ALBETE). The ALBETE strategy eliminates unreliable parts of the classification boundary first, and then generates more training data upon the more reliable classification boundary.

To demonstrate the performance of the proposed strategy, an empirical active learning algorithm named *the ALBETE with network inversion* (ALBETE/NI) is proposed as a practical implementation of the strategy. In the ALBETE/NI, detection of input vectors lying upon the classification boundary is regarded as finding candidates for new training data. To find those candidates, the ALBETE/NI uses an gradient descent algorithm called the *network inversion* (NI) algorithm, which can detect input vectors lying on the classification boundary without using any information except the input-output mapping of the MLP.

The ALBETE/NI uses a candidate detected by the NI as a new training datum only if there is no training datum within a certain threshold γ from the candidate. This indicates that the ALBETE/NI generates new training data only upon parts of the classification boundary considered unreliable under γ. As a result, the ALBETE/NI can preferentially eliminate those unreliable parts, and make the whole classification boundary reliable under γ. At the same time, the ALBETE/NI can avoid excessively generating training data far from the decision boundary, which could become redundant in the future, because the density of newly generated training data is limited by γ. When the whole boundary becomes reliable under the current γ, the ALBETE/NI decreases γ and then generates more training data upon the more reliable classification boundary. As a result, the ALBETE/NI can generate many training data only nearby the decision boundary even through the location of the boundary is unknown, and allow MLPs to achieve accurate classification with fewer training data.

The experimental results in this chapter show that an MLP with the ALBETE/NI can achieve accurate classification with fewer training data than that for some typical active learning algorithms. Moreover, it does not require expensive computations. Accordingly, the ALBETE strategy can effectively construct appropriate training data sets with less redundancy.

3 Acceleration Techniques for Network Inversion

Like the *backpropagation* (BP) algorithm, which is the most popular algorithm for adjusting weights of an MLP, the NI also minimizes the squared error between the desired output values and actual output values of an MLP defined by,

$$E(\boldsymbol{x}) = (y(\boldsymbol{x}) - \tau)^2,$$

where $y(\boldsymbol{x})$ denotes the output value of an MLP for input vector $\boldsymbol{x}$, and τ is the output value of the MLP for input vectors lying upon the classification boundary. Although the BP algorithm minimizes the squared error by adjusting weights of an MLP, the NI minimizes it by changing input values for an MLP:

$$\boldsymbol{x} := \boldsymbol{x} - \epsilon \cdot \nabla E(\boldsymbol{x}),$$

where ϵ denotes a small positive constant to adjust the amount of updating vector $\boldsymbol{x}$. Vector $\boldsymbol{x}$ arrives at the classification boundary when the squared error becomes zero.

This chapter discusses the computational cost of finding candidates by the NI. First, this chapter points out the reason why the NI can be time-consuming, and shows that the *gradient-size* of an input-output mapping of an MLP has a harmful influence on the computational cost of the NI. Next, by taking advantage of the duality between weights and input values in minimizing the squared error, some acceleration techniques for the BP algorithm are applied to acceleration of the NI in order to eliminate the harmful influence of the gradient-size. Finally, the effectiveness of the acceleration techniques is discussed through two simple experiments.

The experimental results in this chapter show that one of the acceleration techniques called the resilient backpropagation algorithm can greatly accelerate the detection of the candidates because it has been developed to eliminate the harmful influence of the gradient-size. This acceleration can reduce the computational cost of the ALBETE/NI required for construction of an appropriate training data set.

4 Active Learning using Self-organizing Neural Networks

There are many classification problems where inputs for an MLP are considered to be randomly drawn from an input space according to *an input distribution*, which is a probability distribution of inputs in the classification problem being learnt. In the case where the classification accuracy of an MLP is evaluated over an input distribution, the ALBETE/NI cannot directly improve the accuracy because it cannot take the distribution into account.

In this chapter, another implementation of the ALBETE strategy named *the ALBETE with topology-preserving maps* (ALBETE/TPM) is proposed to improve the classification accuracy of MLPs with taking account of an input distribution. In the ALBETE/TPM, an unsupervised learning neural network model called a self-organizing neural network (SONN) is iteratively applied to representation of an input distribution. When the representation is formed by an SONN, an input distribution is represented by the distribution of *node vectors* of the SONN. At the same time, an SONN can also form a topology-preserving map, where adjacent regions in an input space have to correspond to adjacent node vectors and vice versa. Topology-preserving maps formed by SONNs can provide a useful information for detection of candidates for new training data, which exist upon the classification boundary of an MLP. Furthermore, they are also useful to evaluate the reliability of the classification boundary with taking an input distribution into account. Therefore, by using those topology-preserving maps, the ALBETE/TPM can effectively improve the classification accuracy of MLPs evaluated over an input distribution.

The experimental results in this chapter show that, with fewer training data, the ALBETE/TPM can allow MLPs to achieve accurate classification evaluated over an input distribution. Moreover, the ALBETE/TPM can make classification of an MLP more accurate than that for passive learning in the case where the network size of the MLP is fixed. Consequently, the ALBETE/TPM demonstrates that the ALBETE strategy can effectively provide only useful training data for MLPs to achieve accurate classification, even in the case where the classification accuracy is evaluated over an input distribution.

In addition, to quickly represent an input distribution, this chapter proposes an unsupervised learning algorithm named *the law of the jungle* (LOJ). A variation of the LOJ named *the LOJ with competitive Hebbian rule* (LOJ/CH) is also proposed for SONNs to rapidly form a topology-preserving map used by the ALBETE/TPM. The experimental results for the LOJ show that SONNs with it can represents input distributions more rapidly than those with Kohonen learning, which is the most popular unsupervised learning algorithm. These results indicate that the ALBETE/TPM can easily use a lot of topology-preserving maps formed by using the LOJ/CH.

5 Conclusions

Main results in this thesis are summarized in this chapter. In conclusion, the avoidance of generating temporarily useful training data makes active learning more useful for construction of appropriate training data sets. The following gives future research works:

- Application of the ALBETE strategy to more realistic problems.
- Analysis of the performance of the ALBETE/TPM in detail.
- Study on active learning applicable not only to classification problems, but also to the other problems such as feature extraction, prediction, and non-linear control systems.

論文審査の結果の要旨

パターン分類問題に広く応用されているニューラルネットワークは，入力パターンとそれに対する正しい分類結果から構成される訓練データを用いて，分類問題を学習できる．一般に訓練データの生成は高価であるため，高い分類精度を実現できる必要最小限の訓練データの生成が望まれている．

著者は，この点に着目し，分類精度向上に貢献する学習データのみを用いて訓練を行なう能動的学習の戦略を提案し，本戦略に基づいた訓練データ生成手法を研究・開発した．本論文はその成果をまとめたもので，全編5章よりなる．

第1章は緒論である．

第2章では，既に存在する訓練データの密度を考慮することで，分類精度向上に有効な訓練データを段階的に生成する訓練データ生成戦略 “ALBETE” を提案している．本章ではさらに，入力分布が一様である分類問題に対し，提案した戦略をネットワーク逆写像法を用いて具現化し，低コストで高効率な訓練データ生成が可能であることを数値実験により明らかにしている．また現実問題への応用例として顔画像認識問題に提案戦略を適用し，認識率を効果的に向上させる訓練データの生成が可能であることを明らかにしている．これらの成果は，分類能力向上に必要な訓練データの生成コストを著しく軽減するという観点から高く評価できる．

第3章では，ネットワーク逆写像法と誤差伝搬法の類似性に着目し，誤差逆伝搬法の高速化手法をネットワーク逆写像法に導入することにより，訓練データ生成に必要な計算コストを削減する手法を提案している．数値実験の結果，提案手法が従来の手法に比べて，極めて少ない繰り返し回数で訓練データを探索可能であることを示している．このことは，特に大規模な分類問題を扱うときに効果的であり，実用上有用な成果である．

第4章では，入力分布が一様ではない分類問題に対して，ALBETE 戦略に基づく訓練データ生成手法について検討している．本章ではまず，入力分布を高速に推定するために，自己組織化ニューラルネットワークの新しい学習手法を提案している．一般に，自己組織化ニューラルネットワークは有限個のノードによって分布を推定するが，本学習手法では，推定に不要なノードを推定に有用なノード近傍に直接移動させることにより，入力分布の推定を高速に行なうことが可能である．次に，推定された入力分布を利用して ALBETE 戦略を具現化する手法を提案している．そして，本手法により，入力分布が一様でない分類問題に対しても，従来の手法と比べて著しく少ない訓練データで高精度の分類能力を実現できることを数値実験を通して明らかにしている．これは重要な成果である．

第5章は結論である．

以上要するに本論文は，ニューラルネットワークの能動的学習の効率を高めることで，訓練データの生成という高価な作業を軽減する手法を考案し，その理論と応用方法を確立したもので，情報基礎科学の発展に寄与するところが少なくない．

よって，本論文は博士（情報科学）の学位論文として合格と認める．